2000

Something from nothing: Augmenting a paper-based work practice via multimodal interaction

David R. McGee, Philip R. Cohen, and Lizhong Wu
Center for Human Computer Communication
Department of Computer Science and Engineering
Oregon Graduate Institute
Portland, OR 97006 USA
+1 503 748 1602
{dmcgee, pcohen, lwu}@cse.ogi.edu
http://www.cse.ogi.edu/CHCC/

ABSTRACT
In this paper, we describe *Rasa*: an environment designed to augment, rather than replace, the work habits of its users. These work habits include drawing on Post-it™ notes using a symbolic language. Rasa observes and understands this language, assigning meaning simultaneously to objects in both the physical and virtual worlds. With Rasa, users rollout a paper map, register it, and move the augmented objects from one place to another on it. Once an object is augmented, users can modify the meaning represented by it, ask questions about that representation, view it in virtual reality, or give directions to it, all with speech and gestures. We examine the way Rasa uses language to augment objects, and compare it with prior methods, arguing that language is a more visible, flexible, and comprehensible method for creating augmentations than other approaches.

Keywords
Phicons, ubiquitous computing, augmented reality, mixed reality, multimodal interfaces, tangible interfaces, invisible interfaces

INTRODUCTION
When people communicate, they often use found objects—they scribble on napkins, draw in the sand, move salt and peppershakers, etc. These objects come to serve as temporary or semi-permanent containers of meaning in virtue of linguistically based multimodal interactions. It is with language that we are able to create something from nothing—to imbue a *tabula rasa* with meaning. This paper takes a first step towards building augmented environments that offer such flexibility. The goal is to invisibly augment a real environment to support existing work practices, and to extend to this environment the benefits of digitization.

In the Proceedings of the *ACM Designing Augmented Reality Environments (DARE'2000)*, April 12-14, Helsinor, Denmark, 71-80.

Work Practice
The work practice that we augment is that of a military command post. Field studies at the Marine Corps bases at Twenty-nine Palms and Camp Pendleton, California, were conducted, during which commanders and their subordinates were observed engaging in field training exercises. Spoken and gestural interactions in the command post were videotaped and transcribed.

One responsibility of officers is to track the movement of friend, foe, and neutral parties. To do this, they construct a kit of useful items from everyday objects. These kits always include a high-fidelity paper map of the terrain, some way to hang the map, objects that are used to represent the various forces, and pens. One of the objects most often chosen to represent forces on the map is a 3M Post-it™ note.

Although the advantages of paper as a tool are numerous, it is worth reiterating a few that have particular relevance to this task. Paper is high in resolution, cheap, lightweight, malleable, and can be rolled up and taken anywhere. Indeed, command posts will often roll up their maps with attached Post-its, move to a new location, unroll, and within minutes continue operations as before.

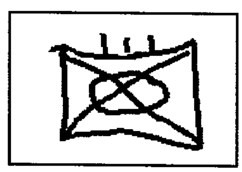

Figure 1. Typical unit symbol

When unit positions are reported over the radio, an officer draws a symbol depicting the unit's strength, composition, and other relevant features in ink on the Post-it (e.g., Figure 1). The unit symbol is one among thousands derivable from a composable *language* for these pictograms that is learned during officer training and used daily. The location of the Post-it on the map represents the unit's position in the real world (Figure 2). Somewhere between several dozen and several hundred of these Post-its may be arrayed on a typical command post map. In addition to the information represented by the Post-it Notes on the map, auxiliary information is available on nearby charts. At any time, anyone

Figure 2. Actual map, overlay, and Post-its in Marine Corp command post

techniques described above. The officers synchronize the paper and digital copies of the current situation in order to mitigate the risk of losing command capability should the computing system fail.

One way to reduce the workload and training for this task is to merge the old with the new—to use the process of augmenting physical icons like the Post-it Notes as the source of meaning for the digital domain.

OBJECTIVE

The objective of this paper is to examine how to augment a successful work practice without replacing the tools of the trade. In the next section, we describe our approach to augmenting physical objects. Examples of Rasa in use and its architecture are presented. We then discuss related work, focusing primarily on how Rasa's use of language to augment objects distinguishes it from its predecessors.

who looks at the map should have a clear picture of the current state of affairs.

In response to radio reports, it is the job of users in each command post to keep all of this information as accurate and as complete as possible, so that their superiors can make critical decisions efficiently and quickly. The primary reason for digitizing this work practice has been to facilitate and improve communication efficiency up the organizational hierarchy.

However, despite major efforts to digitize command and control for ground forces, command posts are still very much a paper, acetate, and grease pencil affair. Command posts must be absolutely robust to all kinds of failure (e.g., hardware, software, communications, and power). Because they are subjected to oppressive environmental and operational conditions, these types of failures are common. During our recent observations, communications were intermittent, power generators failed, and the desert conditions proved fatal to hardened desktop computers. Humans are another strained resource in this environment—the workers there are heavily task loaded. Overall, these conditions lead to a lack of tolerance for any human interface that is confusing, unforgiving, or difficult to operate.

Indeed, the effort spent on this work practice has doubled as command posts have begun digitizing this task. Not only does the officer, or a computer specialist assigned to her, update the unit's position on her graphical user interface (GUI), she continues to update an acetate map overlay hidden behind the projector's screen using the Post-it

APPROACH

Most researchers in augmented environments cite the gulf between human and machine, between work practice and computing environment, as *the* motivating problem. However, successes in augmented environments have been crafted primarily on the computational side of that divide.[1] One researcher who attempted to bridge that gulf by building from the other side was Krueger. In describing the human-machine interface and its design, Krueger argued that: "The computer should adapt to the human, rather than the human adapting to the computer [13]." His goal was "to create unencumbering, environmental artificial realities." As a scientist interested in applying these concepts in artistic ways, however, Krueger did not seek to build any of his responsive environments based around a specific work practice,[2] as we do.

[1] Ishii's Urp [24] is a notable exception.

[2] His VideoDesk prototype is as close as he came to building upon a work practice.

Before we can hope to build a responsive augmented environment where the users remain unencumbered and unattached to computer displays, we must first define what it means to augment the environment. With Rasa, we seek to augment the tools in a work practice, and to do so such that neither the tools nor their use is significantly altered. By "augmenting," we mean

> *Adding something to a real world object to cause it to represent, denote, or be associated with something else.*

Thus, we will speak synonymously of representational and denotational augmentations, as well as associational augmentations. For this task, users are already augmenting paper—creating denotation relationships between Post-its and the things they represent by drawing glyphs in a symbolic language on each note (see Figure 3). If a *system* could perceive these augmentations by understanding this language,[3] then the user could continue to employ familiar tools, which would then be coupled to the digital world.

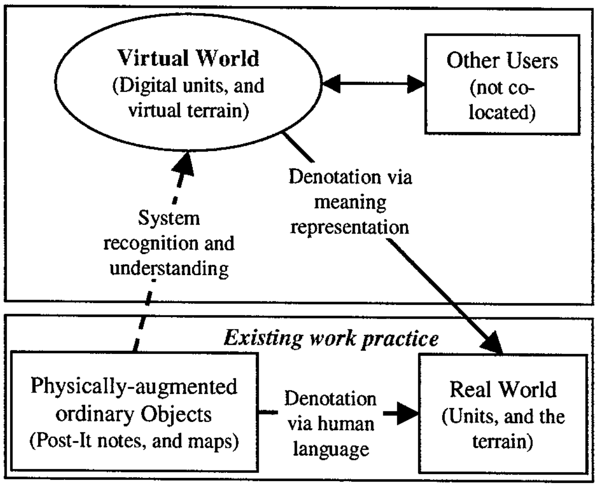

Figure 3. Using augmentations in the real world to produce meaning in the virtual world

To accomplish this, our system should recognize the users' symbology, as well as their handwritten and spoken language. From this multimodal language, the system should derive a denotation relationship from the note to a particular virtual object in the digital world, which itself denotes a real world entity.[4] Because the system supports augmentation through understanding the user's language, the user need not even know that objects in his work practice have been further analyzed by a system, or even that a computer system is operating behind the scenes.

From these insights and the nature of the work practice, we identified five key constraints for Rasa's design, specified in Table 1.

Table 1. Rasa's design constraints

Minimality Constraint	Changes to the work practice must be minimal. The system should work with user's current tools, language, and conventions.
Human Performance Constraint	Multiple end-users must be able to perform augmentations.
Malleability Constraint	Because users gain information about the real world object over time, the meaning of an augmentation should be changeable; at a minimum, it should be incrementally so.
Human Understanding Constraint	The users must be able to perceive and understand their own augmentations unaided by technology. Moreover, multiple users should be able to do likewise, even if neither are spatially nor temporally co-present. Users must also understand what the augmentation entails about the corresponding objects in the real world.
Robustness Constraint	The work must be able to continue without interruption should the system fail.

Two derived constraints immediately follow from these: First, a corollary of the minimality, human performance, and human understanding constraints is that in order to function in the given environment, human-machine interfaces must be based on the current work style, yet be invisible, including those interfaces necessary to augment an object or change the meaning of an augmentation.

A second consequence, based on the minimality and human understanding constraints, is that the system must rely on the language of the work practice to establish the proper representational relationships between the augmented objects and the digital world. Those denotational relationships should be analogous to the ones being created between the Post-its and the real world entities that they represent. Of course, it is the job of the system's semantic interpreter to ensure that these relationships are consistent.

ILLUSTRATION

When the user first sets up his station in the command center, he unrolls his map and attaches it to a SmartBoard or other touch-sensitive surface (see Figure 4). The paper map can now be registered to a position in the world by tapping at two points on the map and speaking the coordinates for each point. Immediately, Rasa is capable of

[3] By "language," we mean an arrangement of perceptible "tokens" that have both structure and meaning to their users. This definition is meant to subsume both natural spoken and written languages, as well as diagrammatic languages such as military symbology.

[4] Precisely how the digital entity comes to have the intended meaning to the user is a complex issue that is beyond the scope of this paper.

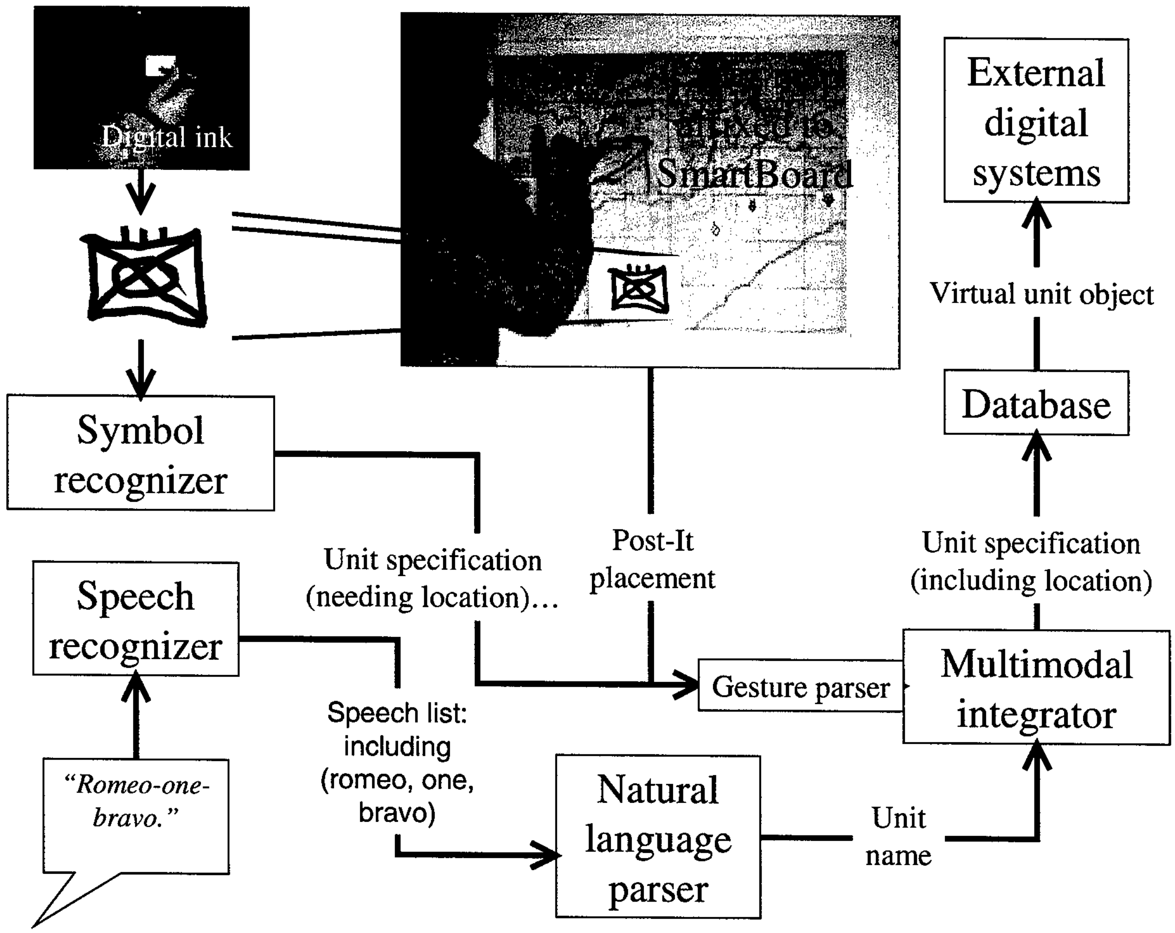

Figure 4. Rasa system architecture

projecting information on the paper map, or some other display, from its digital data sources. For example, Rasa can project unit symbology, other map annotations, 3D models, answers to questions, etc.

When he receives a report of a new unit, he draws the appropriate symbol on a Post-it. At the same time, he can modify the object using spoken language that would otherwise be difficult to capture with the symbology. These are not voice annotations, but actual instructions to further transform the meaning of the augmented object.[5] For instance, he draws the unit symbol in Figure 1 and at the same time says the unit's name, for example "*Romeo-One-Bravo*." Next, the user places the Post-it on a registered map of the terrain, in response to which Rasa says, "*Confirm: Enemy mechanized regiment called Romeo-one-bravo has been sighted at nine-six, nine-four*[6]." The user can then disconfirm the system's response if it is in error. There is no need to confirm the command, if it is correct, since Rasa understands implicit confirmations [14].

The user can further augment the note with speech, gesture, or both. Drawing an arrow starting near the center of the Post-it note he says, "*Romeo-one-bravo is moving in this direction at 20 kilometers per hour*." Rasa projects this new fact onto the paper map in the form of an arrow labeled "*20 kph*." In order to get a better look, the user decides to explore a 3D representation of the terrain. He points once again at Romeo-one-bravo and says, "*Fly me to this regiment*." A 3D fly-through of the terrain appears, projected onto a nearby flat surface.

IMPLEMENTATION

Rasa is comprised of a set of agents—autonomous software components that communicate using an agent communication language in the Open Agent Architecture (OAA) [6]. Because of this, Rasa can take advantage of the QuickSet multimodal system [7]. QuickSet lets users create objects such as military units and control measures (points, lines, and areas) on a map through use of speech and gesture. Rasa uses QuickSet's database and text-to-speech agent, as well as its speech and gesture recognizer

[5] Rasa objects could contain voice annotations as well.

[6] Nine-six, nine-four is a local coordinate in the military grid reference system.

agents. Rasa goes further by extending QuickSet's language and enhancing its user interface, natural language parser, and multimodal integrator. Rasa also adds a new military symbology recognizer. The agents are described in the subsections below.

Gesture Recognizers

Rasa incorporates the existing QuickSet gesture recognizer agent to recognize simple symbolic gestures, such as points, lines, and arrows. In addition to this agent, we have developed a recognizer for military unit symbology. Both recognizers are based on a hierarchical recognition technique called Member-Team-Committee (MTC) [28].

The MTC optimizes pattern recognition robustness by weighting the contributions derived from individual recognizers based on their empirically derived relative reliabilities. It is comprised of a three-tiered divide-and-conquer architecture with multiple members, multiple teams, and a committee. The members produce local posterior estimates, and then report their results to one or more "team" leaders, which apply weighting parameters to the scores. Finally, the committee weights the results of the various teams, and reports the final results.

Recognizing handwritten military unit symbols is difficult, because:

1. The stroke patterns and sequences are not stable across users, and
2. A large image size is required in order to fully represent and distinguish different symbols.

However, the MTC technique is able to overcome these difficulties, recognizing 200 different military unit symbols, and achieving a better than 90% recognition rate. The hierarchical approach appears particularly effective at handling data modeling problems involving high-dimensional input features typical of gestural pattern recognition.

User Interfaces

Use of everyday objects distinguishes Rasa's interaction style from that of QuickSet. To support everyday objects, the QuickSet user interface was modified for Rasa so that it becomes a virtual interaction surface underneath the paper map. Whenever the user touches the map and the touch-sensitive surface beneath it, they, are in fact, interacting with Rasa's user interface. In addition to the Post-its, entities and other map annotations can be created multimodally and then projected onto the paper map. For instance, if hazardous terrain exists, the user could add appropriate symbology for restricting access to that area. He does this by sketching a circular gesture on the map with his finger, or to ensure a semi-permanent annotation remains, he can choose to draw with a pen on the map's transparent plastic overlay, while saying, "*No go area.*" The resulting military icon for that object is projected onto the map.

A pad of Post-its is affixed to a Cross Computing iPen Pro™ digital pen tablet, which captures the "ink" digitally, while the pen produces real ink on each note. The interface runs on the system connected to the iPen Pro tablet, but there is no computer or user interface visible, other than the Post-its themselves. The fact that drawing on the note results in the capture of digital ink is invisible to the user.

Speech recognition and text-to-speech

The TTS agent provides spoken feedback whenever visual feedback is infeasible. For a system composed of several "invisible" interfaces, this occurs quite frequently. The speech recognition agent uses Dragon Systems or Microsoft's SAPI speech recognition engines, or IBM's Voice Type Application Factory (VTAF); all are continuous, speaker-independent recognizers, however, both Microsoft and Dragon Systems' recognizers can be trained to better support individual users and working environments. The SAPI-compliant engines use a context-free grammar, while the VTAF engine uses a bigram model. The vocabulary is approximately 675 words, and the grammar specifies a far greater number of valid phrases. Both the text-to-speech and the speech recognition agents were developed for use with the QuickSet system, but Rasa is our first agent-based application to take substantial advantage of TTS.

Natural language and gestural parsers

The natural language agent employs a definite-clause grammar and produces a meaning representation in the form of typed feature structures [10]. Currently, for this task, the language consists of map-registration predicates, noun phrases that label entities, adverbial and prepositional phrases that supply additional information about the entity, and a variety of imperative constructs for supplying behavior to those entities or to control various systems.

The gestural parser is a subsystem of the multimodal integrator. Its job is to produce meaning representations from a basic list of recognition hypotheses and probability estimates from the gesture recognizer. In general, parsers can create multiple interpretations for each recognition hypothesis. For example, an enclosing line gesture (one recognition hypothesis) has at least two meaningful interpretations—a selection is being made or an area is being created. It is the job of the multimodal integrator to determine which interpretation makes the most sense. In this case, the integrator is assisted by the user interface and natural language interpreters. The user interface estimates which displayed objects are associated with any particular ink stroke (i.e., which objects are touched, encircled, etc.). The natural language interpreter can further disambiguate whether a selection action is occurring or an area is being created. If an object that is in the list the user interface provided is referred to in the spoken language, the multimodal integrator can judge

that as strong evidence in favor of an action being taken upon that object.

Multimodal Integration

QuickSet's multimodal integration technology uses declarative rules to describe how the meanings of input from speech, gesture, or other modalities combine. Precursors to this fusion architecture include the original "Put-That-There" [1], and other approaches [5, 12, 16, 18]. What distinguishes QuickSet's multimodal integration architecture, however, from other approaches is that it:

1. supports expressive (i.e., beyond pointing) gestural components as first-class citizens in the fusion of modalities
2. accommodates multimodal expressions that include multiple spoken or gestural components in a single utterance
3. provides a well-understood, generally applicable common meaning representation (feature structures) for the different modes and a formally well-defined mechanism (unification) for multimodal integration

The basic algorithm underlying multimodal integration is feature structure (or "frame") unification [2, 3], a generalization of term unification as found in logic programming languages. Unification is appropriate as the basic information-fusion operation because it combines both complementary and redundant information, but rules out incompatible attribute values.

One of the basic rules in the QuickSet multimodal integrator is used to fuse any semantically complete utterance that is only lacking a location feature with a gesture that provides a compatible location feature. Unification ensures us that both inputs contain compatible features (attribute/values). A declarative set of temporal constraints is used that were developed based on empirical investigation of multimodal synchronization [20]. Spatial constraints are used for combining gestural inputs, and new constraints can be declared and applied in any rule.

In general, multimodal inputs are recognized, and then parsed, producing meaning descriptions in the form of typed feature structures. The integrator fuses these meanings together by evaluating any available integration rules for the type of input received and those partial inputs waiting in an integration buffer. Compatible types will be unified, and then constraints will be satisfied. Successful unification and constraint satisfaction results in new merged feature structures. The highest ranked semantically complete feature structure is executed. If none are complete, they wait in the buffer for further fusion. A complete description of the integration architecture can be found in [10, 11].

Functional Description

An example of how Rasa functions, as illustrated in Figure 4, is described below. When the user draws a unit symbol on a Post-it Note, the user interface stores the ink for potential image analysis, activates the speech recognizer, and sends ink to the gesture recognizer. Immediately, both the speech recognizer and the gesture recognizers begin analyzing their respective inputs and send any interpretations to their respective parsers via the OAA. Each parser produces a set of possible meaning representations that are routed to the multimodal integrator. The integrator unifies these interpretations using a statistical algorithm to determine the best joint interpretation, given a set of empirically based multimodal constraints [20]. The unit's unified meaning representation, however, still lacks a location component.

After drawing the symbol, the user picks up and places the Post-it Note on a registered map. Rasa integrates the new coordinate information with the recently unified unit interpretation and requests the database agent to perform an insert operation on the new unit, creating a unit at that location. Whenever the unit is reported to have moved, the user simply picks up the Post-it, and moves it to the new location on his map. After the architecture processes those multimodal inputs, in a fashion similar to that described above, the database receives an update request for the unit being moved; it then modifies the unit's location appropriately.

In summary, the Rasa system augments this environment with invisible interfaces, leaving the original work practice intact. Two personal computers, their monitors, etc., added to support this augmentation, are hidden from view. A SmartBoard or similar touch-sensitive board, an ink-producing digital pen, like the iPen Pro from Cross Computing, and a wireless microphone or microphone array are not hidden, but remain unobtrusive. Based on technology developed first in QuickSet, Rasa transforms multimodal input in the form of speech, gestures, and the manipulation of real world objects into meaning representations, and from meaning representations into commands to a variety of distributed agents that communicate using the OAA.

As a descendant of QuickSet, and specifically as an interface in a domain that has been studied extensively, map-based interactions, it inherits all of the advantages that can be expected from multimodal interaction in that domain, including, but not limited to,

- User preference
- Human performance advantages
- Fewer errors

as compared with both GUI and speech-only interfaces [8, 19].

RELATED WORK

This work was inspired by visions of ubiquitous computing and augmented reality [13, 17, 26, 27], though our work most closely resembles the recent approaches of Ishii and his students [9, 22-24]. In this section, we compare Rasa to similar research and discuss how prior systems fare with respect to the five design constraints of Table 1.

The Urp [24] system, like Rasa, augments a natural, non-digital work setting. With Urp, planners use building models, rulers, clocks, and other physical objects to design an urban environment. Objects are tagged by patterns of colored dots, and if a pattern is recognized, the vision system sends Urp the associated object's location. Both Urp and Rasa use tools that are natural and familiar in their setting. With Urp, augmented objects "behave" as you would expect them to: rulers measure distances, clocks mark time, and so on. The object's physical characteristics and the environment it inhabits govern these expectations. With Rasa, however, objects are transformed when they are augmented: Post-its come to represent units, whereas before they had no prior meaning in the work setting. This enables Rasa to have as many augmented objects as pieces of paper, whereas Urp and systems that rely on the physical properties of objects to denote specific meaning will likely have a smaller number of augmented objects whose meanings are fixed in advance by the developer. As such, the user cannot easily change them (cf., the *malleability constraint*).

Three projects that have augmented paper are most relevant to this research. The DigitalDesk [27] augments office work by introducing paper into a workstation environment. Through computer vision, users can point at numbers on a real piece of paper, in response to which the system performs optical character recognition and pastes the recognized number into the system's calculator. Similarly, regions of real paper, like a sketch on a napkin, can be cut and pasted into a painting program. The transBOARD [9], a shared whiteboard, uses barcode-tagged cards to hold digital ink. However, the ink can only be retrieved when scanned by a barcode reader connected to a desktop computer. The Intelligent Room [4] uses Post-it Notes to activate behaviors in the room. Different colored Post-it Notes are used so that they can be easily distinguished from each other and from the supporting table by a vision system. Ink on each note is used only to remind the user of the action to expect, not as input to the system.

None of these paper-based systems claims to support a formalized work practice, nor do they attempt to augment a pre-existing situation, as does Rasa. Moreover, none of the aforementioned systems interprets the human act of augmenting paper in order to create a digital representation, which is our *human performance constraint.*

Both the Passage system [21] and the recent RFID (radio frequency identifier) research at Xerox [25] meet this constraint and provide some flexibility in changing that data—the *malleability constraint.* Within the Passage concept, meaning can be linked graphically to a physical object whenever that object is placed on a "bridge." In the initial prototype, the bridge is a scale and recognizes objects based on their weight. With the RFID system, tags are hidden in books, documents, watches, etc. As with Passage, associational augmentations can be formed when the tags are first detected. Unlike Rasa, however, these systems do not yet support a pre-existing work practice, nor can users learn what information is associated with an object unless the users and the object are adjacent to a bridge or detector. More generally, associational augmentation methods like these and others, such as the use of colored dots, glyphs, or bar codes, fail to present the linked digital information to the user without the assistance of technology. Thus, these methods would not satisfy our *human understanding constraint.*

It is because users of Rasa are augmenting objects with written language rather than simply associating physical objects with digital information that these augmentations remain both visible and understood. Furthermore, with written language, additional content can be added to an augmented object, thereby recording a history of changes to the augmentation that remains permanent and visible. This particular aspect of *malleability*, incrementality with permanence, does not hold for all modalities of language. In particular, speech does not have this property. However, the Post-its in the command post are currently augmented with speech when the information being added tends to be transitory. These invisible adjustments to the meanings are shared with other users when necessary, or when questions regarding the objects arise. Furthermore, Rasa has the option of making these changes visible, since it also maintains a comprehensive representation of the augmented object. For example, it can project any entity's transitory properties onto the paper map if need be.

Finally, according to the *robustness constraint*, the augmented environment must allow users to continue to work even in the face of a power or other type of failure. Because Rasa augments an existing paper-based work practice, users could grab a flashlight, a ballpoint pen, and a Post-it Note, and continue working. After the system restarts, even if Rasa lost all data, the information can be recovered with language common to the work practice. A user could point at a unit, read the symbology stored permanently on the Post-it Note, and recreate the augmentation with speech. For example, *"This is a friendly medical company called Tango-five-two."*

The Collaborage concept [15], which characterizes augmented systems consisting of a board and various tagged physical information items, has been applied to build

several prototypes at Xerox/PARC. One of these prototypes is an In/Out board system that satisfies many of the constraints described here. With the In/Out board, glyph-tagged magnetized photos can be slid from the Out column to the In column and vice-versa. Within seconds, a vision system recognizes the change in location of the glyph and an In/Out web page is updated to reflect the change in status. If the system were to fail, individuals could still check the physical In/Out board, move their picture from one column to the other, add hand-written annotations, and walk away with up-to-date information. Because objects are augmented using glyphs rather than a natural language, users cannot easily add new digital information to the board. For example, a new employee cannot use any magnet and photograph and expect it to work in the Collaborage.

Since the other augmented environments discussed here rely heavily on computers and computer interfaces in order for the users to *understand* the augmentations and to *use* augmented objects, if any of those systems were to fail, the work would stop.

Limitations of Rasa

Rasa itself has several limitations, including an incomplete vocabulary and grammar. In order to cover the language of this work practice more adequately, data will need to be collected and experiments will need to be run. To complete the circle of observation, engineering, and experimentation, this spring we are scheduled to evaluate Rasa in the field as part of a series of experiments conducted for the Command Post of the Future project at DARPA.

Currently Rasa's users are limited in two ways. A Post-it must be created, and then applied to the map, before a second one is created. Likewise, notes can only be moved one at a time once they are placed on the map. These limitations can be overcome with a minor adjustment to the existing work practice, where spoken interaction is not necessary to use Rasa. While applying the symbology to the Post-it Note, the user would also name the unit. The name can then be used to disambiguate among the potential units if the name is spoken while placing or repositioning the Post-it on the map.

Second, though the entire computing infrastructure in Rasa is essentially hidden from view, it still relies on the SmartBoard. Our colleagues in the military have told us that "a map with a hole in it is still a map, but a computer display (or SmartBoard) with a hole in it, is a rock." Our aim is to extend the existing system with vision components that will eliminate the need for a SmartBoard, and expand the interaction styles available. Using the digital ink stored by the user interface agent, we can use vision to track the location of the digital ink's physical counterpart on the map.

DISCUSSION

Augmentations can be visible or invisible, transitory or permanent. They can describe, denote, or associate one object with another. The work practice for which Rasa was designed requires that the augmentations be visible, while the interface that creates them should be invisible. By augmenting the existing practice and tools, the users themselves create two denotation relationships simultaneously, one connecting a Post-it Note to an entity in the real world, and the other connecting that same note to a virtual entity in a virtual world. It is because of Rasa's ability to recognize and understand the note's augmentations that we can introduce Rasa into that work practice essentially unchanged.

There are several benefits to using language as a method of augmentation. Written language is visible, semi-permanent, and immediately comprehensible. Conversely, spoken language is invisible and transitory (i.e., it is as permanent as the user's memory for recalling the denotation), though it offers a compact, familiar, and efficient way to change meaning representations.

There may well be benefits obtainable from reducing the duplicative and error-prone effort to keep paper and computer systems synchronized, and from eliminating training where it is no longer necessary. Likewise, there may be substantial benefits from digitizing the information. Once entities are added to a common database, many capabilities are enabled. For example, with Rasa, creation of entities on the terrain supports simulation and visualization. Moreover, users can immediately participate in collaborative interaction, with Rasa projecting virtual objects onto each user's paper map.

FUTURE WORK

The ultimate goal of Rasa is to develop multimodal mechanisms for augmenting and interacting with *arbitrary* objects, and in informal encounters. One day soon, we hope, someone could sit down in a Rasa-enhanced environment, meet with his or her commander about evacuating an embassy, and have the following conversation. Grabbing the Tabasco bottle from her ready-to-eat meal, Lt. Smith says, "Sir, if this is the embassy, and that packet of sugar is the airport, why don't we convoy the UN personnel around this hill (points to *a hamburger*) to get them to the helicopters?" An enhanced Rasa should be able to recognize the augmentations, understand the multimodal language, and match the terrain features and unit locations with digital data sources. It could then simulate the evacuation over lunch, and place the plan into a database for further review.

CONCLUSIONS

In this paper, we have described Rasa, an augmented environment that uses multimodal language to augment objects in the real world and in real time. We showed how Rasa was designed to minimally impact the current

work practices of a military command post, in which users already augment objects, like Post-it Notes, with symbology in order to denote real world objects on the terrain. Rasa understands that symbology, and users' accompanying multimodal (speech/gesture) input, enabling users to create digital representations of the entities with which they are interacting without even knowing that a computer is involved. Finally, comparing it to other schemes, we have shown that language can be a flexible, expressive, and comprehensible way to augment everyday objects.

ACKNOWLEDGEMENTS

Many thanks to the men and women of the Fifth and Eleventh Regiments, First Division, Marine Corps, Col. Steve Fisher, (USMC, retired), and Col. Jack Thorpe (USAF, retired). This work was supported in part by the Information Systems Office of DARPA under contract number N66001-99-D-8503, and also in part by ONR grants: N00014-95-1-1164, N00014-99-1-0377, and N00014-99-1-0380. The views presented here are those of the authors and do not represent those of the US Government.

REFERENCES

1. Bolt, R.A., *"Put-That-There": Voice and gesture at the graphics interface.* Computer Graphics, 1980. 14(3): 262-270.
2. Carpenter, R., Typed feature structures: Inheritance, (in)equality, and extensionality, in the *Proceedings of the ITK Workshop: Inheritance in Natural Language Processing,* Institute for Language Technology and Artificial Intelligence, Tilburg University, Tilburg, 9-18.
3. Carpenter, R., *The logic of typed feature structures.* 1992, Cambridge England: Cambridge University Press.
4. Coen, M.H., Design principles for intelligent environments, in the *Proceedings of the Conference on Artificial Intelligence (AAAI '98),* July 1998, American Association for Artificial Intelligence, 547-554.
5. Cohen, P.R., Integrated interfaces for decision support with simulation, in the *Proceedings of the Winter Simulation Conference*, ACM Press, 1066-1072.
6. Cohen, P.R., Cheyer, A., Wang, M., and Baeg, S.C., An open agent architecture, in the *Proceedings of the AAAI Spring Symposium,* Mar. 1994, Reprinted in *Readings in Agents,* Huhns, M. and Singh, M. (eds.), Morgan Kaufmann Publishers, San Francisco, 1-8.
7. Cohen, P.R., Johnston, M., McGee, D.R., Oviatt, S., Pittman, J., Smith, I., Chen, L., and Clow, J., QuickSet: multimodal interaction for distributed applications, in the *Proceedings of the Fifth Annual ACM International Multimedia Conference (Multimedia '97),* November 1997, ACM Press, 31-40.
8. Cohen, P.R., McGee, D.R., and Clow, J., The efficiency of multimodal interaction, in the *Proceedings of the Applied Natural Language Programming Conference (ANLP'00),* April, 2000, ACL.
9. Ishii, H. and Ullmer, B., Tangible bits: towards seamless interfaces between people, bits and atoms, in the *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '97),* March 1997, ACM Press, 234-241.
10. Johnston, M., Unification-based multimodal parsing, in the *Proceedings of the 17th International Conference on Computational Linguistics and the 36th Annual Meeting of the Association for Computational Linguistics (COLING-ACL 98),* August 98, ACL Press, 624-630.
11. Johnston, M., Cohen, P.R., McGee, D.R., Oviatt, S.L., Pittman, J.A., and Smith, I., Unification-based multimodal integration, in the *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics (ACL '97),* March 1997, ACL Press, 281-288.
12. Koons, D.B., Sparrell, C.J., and Thorisson, K.R., *Integrating simultaneous input from speech, gaze, and hand gestures*, in *Intelligent Multimedia Interfaces*, M.T. Maybury, Editor. 1993, AAAI Press/MIT Press: Cambridge, MA. 257-276.
13. Krueger, M.W., *Artificial Reality II.* 1991, Reading, MA: Addison-Wesley. p. 286.
14. McGee, D.R., Cohen, P.R., and Oviatt, S., Confirmation in multimodal systems, in the *Proceedings of the International Joint Conference of the Association for Computational Linguistics and the International Committee on Computational Linguistics (COLING-ACL '98),* August 1998, ACL Press, 823-829.
15. Moran, T.P., Saund, E., Melle, W.V., Bryll, R., Gujar, A.U., Fishkin, K.P., and Harrison, B.L., The ins and outs of collaborative walls: Demonstrating the Collaborage concept, in the *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '99),* May 15-20, 1999, ACM Press, CHI'99 Extended Abstracts, 192-193.
16. Neal, J.G. and Shapiro, S.C., *Intelligent multi-media interface technology*, in *Intelligent User Interfaces*, J.W. Sullivan and S.W. Tyler, Editors. 1991, ACM Press, Frontier Series, Addison Wesley Publishing Co.: New York, New York. 45-68.
17. Newman, W. and Wellner, P., A desk supporting computer-based interaction with paper documents, in the *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '92),* May 1992, ACM Press, 587-592.

18. Nigay, L. and Coutaz, J., A generic platform for addressing the multimodal challenge, in the *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '95),* May 7-11, 1995, ACM Press, 98-105.

19. Oviatt, S.L., *Multimodal interactive maps: Designing for human performance.* Human-Computer Interaction, 1997. 12 (special issue on "Multimodal interfaces"): 93-129.

20. Oviatt, S.L., DeAngeli, A., and Kuhn, K., Integration and synchronization of input modes during multimodal human-computer interaction, in the *Proceedings of the Conference on Human Factors in Computing Systems (CHI '97),* March 1997, ACM Press, 415-422.

21. Streitz, N.A., Geibler, J., and Holmer, T., Roomware for cooperative buildings: integrated design of architectural spaces and information spaces, in the *Proceedings of the First International Workshop on Cooperative Buildings: Integrating Information, Organization, and Architecture (CoBuild '98),* February 1998, Springer-Verlag, 4-21.

22. Ullmer, B. and Ishii, H., The metaDESK: models and prototypes for tangible user Interfaces, in the *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST '97),* October 1997, ACM Press, 223-232.

23. Ullmer, B. and Ishii, H., mediaBlocks: physical containers, transports, and controls for online media, in the *Proceedings of the 25th International ACM Conference on Computer Graphics and Interactive Techniques,* July 1998, ACM Press, 379-386.

24. Underkoffler, J. and Ishii, H., Urp: a luminous-tangible workbench for urban planning and design, in the *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '99),* May 1999, ACM Press, 386-393.

25. Want, R., Fishkin, K.P., Gujar, A., and Harrison, B.L., Bridging physical and virtual worlds with electronic tags, in the *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '99),* May 1999, ACM Press, 370-377.

26. Weiser, M., *Some computer science issues in ubiquitous computing.* Communications of the ACM, 1993. 36(7): 75-84.

27. Wellner, P., *Interacting with paper on the DigitalDesk.* Communications of the ACM, 1993. 36(7): 87-96.

28. Wu, L., Oviatt, S., and Cohen, P., *Multimodal integration - A statistical view.* IEEE Transactions on Multimedia, 1999. 1(4): 334-341.